Multi-script Baseline Detection Using Perceptive Vision

Aurélie Lemaitre, Jean Camillerapp, Bertrand Coüasnon

Multi-script Baseline Detection Using Perceptive Vision

Aurélie LEMAITRE [a], Jean CAMILLERAPP [b] and Bertrand COÜASNON [b]
[a] *Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes*
[b] *INSA, Avenue des Buttes de Coësmes, F-35043 Rennes*
UMR IRISA, Campus de Beaulieu, F-35042 Rennes
Université Européenne de Bretagne, France
aurelie.lemaitre@irisa.fr

Abstract. Baseline detection is an important step for off-line handwriting recognition of many alphabets. However, it remains a difficult task in the case of varying and irregular handwriting, and sometimes requires script-specific adaptations. In this paper, we propose a new method for baseline detection that is script independent. Indeed, this method is based on the principle of the perceptive vision, which combines several points of view of the same word (from low to high resolution). We validate the genericity of our method by extracting the baseline on three scripts: Arabic, French, and Indian (Bangla), without any adaptation of parameter. We compare our results with other works on the Arabic IFN/ENIT database. Moreover, baseline extraction is often realized independently for each studied word. In this work, we also demonstrate how the use of knowledge on the initial context of the word (text line) significantly improves baseline extraction.

1. Introduction

The baseline is the imaginary line on which leans the writing of a word. Depending on the script, it may be located at the bottom of the writing (Latin alphabet, Arabic) or at the top (Bangla). Baseline detection is an important step for handwriting recognition as the knowledge of the baseline guides character identification and word recognition. This property is used for handwriting recognition of many alphabets such as Arabic, Latin, Indian…

In this paper, we propose to show that the problem of baseline detection can be solved by a generic method, based on the principle of perceptive vision. Thus, we propose to combine a global vision of the words (at low resolution) with a precise analysis at high resolution, in order to find more accurately the position of the baseline. This method can be applied on different scripts (alphabets) without any adaptation of parameter. Moreover, baseline extraction is often realized independently for each studied word. In this work, we show how the use of knowledge on the initial context of the word (text line) can significantly improve baseline extraction.

Our paper is organized as follow. First, we present related works on the two fields of our approach that are the baseline extraction and the perceptive vision. Then, we will present the DMOS-P[1] method that enables the creation of our system based on the perceptive vision. In section 4, we detail our two strategies: one for the baseline extraction in the case of isolated words, and the other one that uses the context of the text line. In the application part, we present results on three kinds of script: Arabic, French, and Bangla writings. These results show that our generic method can be used without any adaptation for three different scripts. They also demonstrate how the knowledge on the context is really important to significantly improve baseline accuracy.

2. Related work

Our work joins two fields: baseline detection in off-line words and perceptive vision in document analysis.

Baseline detection is often considered as a pre-processing treatment and few papers of the literature are only dedicated to this problem. However, it remains an important problem, particularly in the case of handwriting analysis. Concerning Arabic word analysis, the IFN/ENIT base (Abed *et al.* 2007) provides handwritten words and the ground truth of their associated baseline. An overview of the main techniques applied to this base has been recently proposed in (Maddouri *et al.* 2008). However, each proposed method seems particularly adapted to treat one kind of difficulty met in Arabic writing. Moreover, to our knowledge, the proposed methods have not been largely evaluated for other scripts. The baseline extraction is also an important challenge in the case of Bangla handwriting (Bhattacharya *et al.* 2007).

Perceptive vision is the faculty of the human to combine several points of view in order to improve the recognition of an object. This mechanism has been long used in computer vision and more recently in document analysis. For example, it has been used for text line localisation (Déforges *et al.* 1994) or to exploit the properties of the global vision of a structure (Cinque *et al.*1998).

To our knowledge, the perceptive vision has not been used for baseline detection. However, it seems interesting to exploit both the global aspect of the word to obtain rough information on the baseline, and the details of the word to precisely locate the baseline. We had proposed a preliminary work in (Lemaitre *et al.* 2007); we now present our generic system based on perceptive vision for multi-script baseline extraction.

---

[1] Description and MOdification of the Segmentation with Perceptive vision

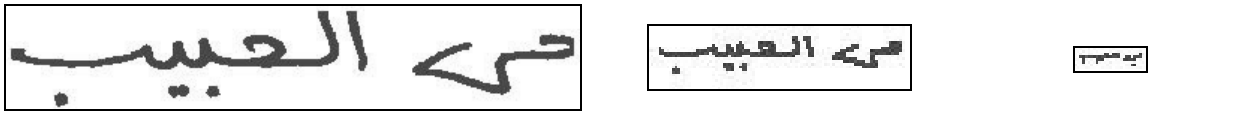
3. Perceptive method DMOS-P

Our system has been realised in the context of a generic method, DMOS-P, which provides a framework for the generation of perceptive cooperation mechanisms. We first recall the principle of DMOS-P method before presenting in details some of its interesting properties for our mechanism of perceptive vision.

DMOS-P method is a generic method for structured document recognition. It is an extension of the DMOS method which has been largely described (Coüasnon 2006). DMOS-P is based on a grammatical formalism, EFP (Enhanced Position Formalism) that makes possible a bi-dimensional description of the physical and the logical structure of a kind of document. Once the grammatical description has been realised, the associated parser is automatically produced by a compilation step. DMOS-P has been used for the analysis of various kinds of documents: incoming mails, tabular structures, old newspapers, archive documents…

In our case, we propose a grammatical description of the baseline of words in a document. This description mainly uses two aspects of DMOS-P perceptive faculties: the cooperation between resolutions and the positioning tool. They have been detailed in (Lemaitre *et al.* 2008); we recall their properties.

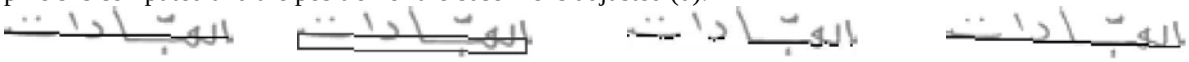
Cooperation between resolutions. One of the particularities of DMOS-P is to enable the cooperation between several points of view (or image resolutions). This combination is a key point of our baseline analysis. We choose to combine three levels of resolutions, built by a down-sampling preceded by a low pass filtering, presented on figure 1. Thus, it makes it possible to combine several perceptions of the word and then to compute the baseline, taking into account the information from each resolution: low resolution gives rough indications on the slope and the thickness of the word, whereas highest resolutions enables a more precise positioning of the baseline.



(a) High resolution (initial image) (b) Medium resolution (Initial image divided by 4) (c) Low resolution (Initial image divided by 16)

Figure 1. Result of the binarization for the three chosen levels of resolution

Positioning tool. In DMOS-P, a specific tool, called *Positioning* enables to adjust the position of a line in a chosen zone, taking into account the presence of black pixels in this zone. More particularly, the positioning can take into account only the lower or upper pixels of vertical run-lengths. This tool is particularly interesting for our work of baseline positioning. An example is presented on figure 2, where we want to adjust the position of a given line at the bottom of a word. Thus, the initial line (a) is used to define a zone of interest (b). In this zone (b), we look for the pixels located at the bottom of vertical run-lengths (c). The average position of these pixels is computed and the position of the baseline is adjusted (d).



(a) Initial line (b) Zone of interest deduced from initial line (c) Selection of bottom pixels in the zone (d) Adjustment of the position of the line on the bottom of the word

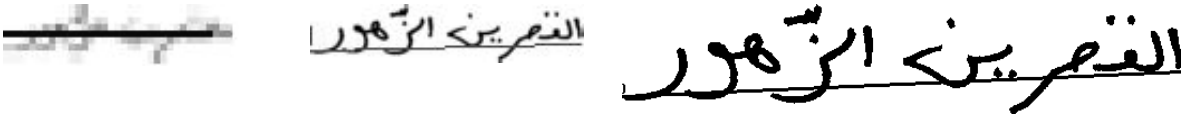
Figure 2. Adjustment of the line position thanks to the positioning tool

4. Baseline extraction thanks to perceptive vision

We now detail the principle of our baseline extraction system, realised thanks to the DMOS-P method. We propose two approaches. On the one hand, we consider the localization of the baseline in isolated words, *i.e.* without context. On the other hand, we study whole pages of text and exploit the properties of each text line to localize the baselines. In both cases, we combine the three levels of perception presented on figure 1, and progressively localize the baseline, from low to high resolution, with more and more precision.

Without context. The principle of recognition of a baseline is (figure 3):

1. At low resolution, estimate the middle part of the word; the result is the line L1.
2. At medium resolution, use positioning tool to adjust L1 on the bottom of the pixels, taking into account the inferior part of the word; the result is the line L2.
3. At high resolution, use positioning tool to adjust L2 on the bottom of the pixels; the result is the final baseline.



(a) Line L1: middle of the word at low resolution (b) Line L2 : bottom of the word at medium resolution (c) Final base line at high resolution

Figure 3. Localization of the baseline without context

Using text line context. We propose a second approach that consists taking into account the initial context of each word in the text line. Thus, having a global vision of the full text line provides more reliable information about the slope and the average thickness of each local word. Consequently, we propose the following process for baseline recognition (figure 4):

1. At low resolution, estimate the position, the slope and the thickness of the global text line. (At low resolution, text lines are detected as line segments.)
2. Apply an algorithm for segmenting lines into words (not detailed here).
3. For each local word, at high resolution, adjust the global line to the extreme pixels thanks to the positioning tool.

We will show on section 5 how the use of the context significantly improves baseline accuracy, particularly for small words or curved text lines where knowledge on the global aspect of the text line is essential (figure 6).

(a) At low resolution, detect text lines as line segments: they give information about slope and thickness.

(b) At high resolution, use the positioning tool to adjust the line seen at low resolution.

Figure 4. Use of context for baseline localization in Bangla script
(In this case, notice that the baseline is at the top of the word)

5. Application on three scripts: Arabic, French and Bangla

Bases. We applied our method on three bases detailed in table 1. The public base IFN/ENIT provides the baseline ground-truths for 6537 words. We manually labelled the baseline ground-truths for a French base provided by RIMES project (Augustin *et al.* 2006) and for a base of Bangla handwritten documents.

Script	Number of words	Number of writers	Nature
Arabic (IFN/ENIT – set a)	6537	102	Isolated words
Bangla	2909	26	Full pages of text
French (RIMES project)	2653	25	Full pages of text

Table 1. Bases of application

Metric. In order to compare our results with existing works, we used the metric proposed with IFN-ENIT basis (Abed *et al.* 2007). This metric estimates the error of a baseline by calculating the average distance between the ground-truth and the proposed result (figure 5). Two thresholds are proposed with this metric: with an average pixel error less than 5 pixels, the baseline is considered as *good*, whereas with an error up to 7 pixels, the baseline is *acceptable* (7 pixels visually seems a small error due to the size of the word). We can apply this metric to our three bases as they have been digitized with comparable dimensions.

Figure 5. Visualization of the baseline error definition

Comparison with other approaches. As we used the metric and the data proposed by IFN/ENIT, we can compare our work with the approaches proposed in the literature (table 2). We can notice that our generic method obtains comparable results with methods that are specifically adapted to Arabic writings.

Baseline quality (average pixel error)	Our method Perceptive vision	(Farooq et al. 2006)	(Maddouri et al. 2008)		
			Skeleton	Primitives	Min-Max
Acceptable (<7)	82.4 %	78.5 %	87.5 %	82.3 %	74.3 %

Table 2. Comparison of the results obtained by our method
with approaches of the literature on the set 'a' of the IFN-ENIT database

A generic method. In order to show the genericity of our method, we applied it on the three bases of table 1. We only choose the positioning parameter: adjust the line on the *lower* pixels for French and Arabic but on the *upper* ones for Bangla script, which main baseline is located at the top of the word. The obtained results are presented in the table 3. The results are promising because they are produced for each script without adaptation of any parameter, contrary to the dedicated methods of the literature.

Baseline quality (average pixel error)	Arabic (IFN/ENIT)	French (RIMES project)	Bangla
Good (<5)	70.4 %	81.0 %	69.7 %
Acceptable (<7)	82.4 %	85.6 %	82.3 %

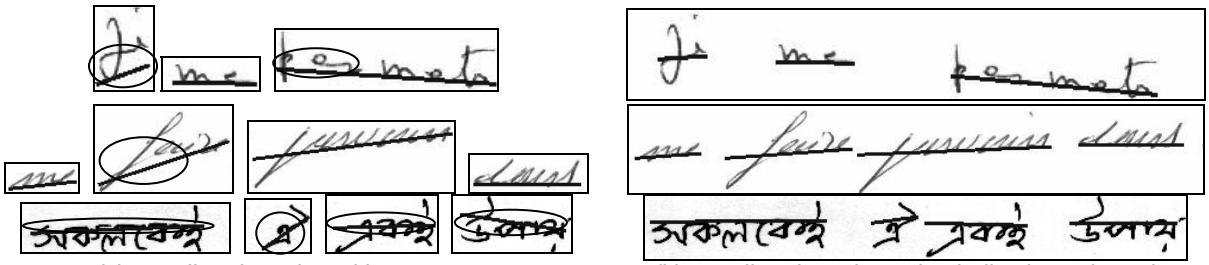
Table 3. Application of our method for isolated words on three kinds of scripts

Use of context improves the recognition. Using the full text line gives better indications on the global position, slope and thickness of the line. Thus, it significantly improves baseline detection. We evaluated this improvement on two bases (table 4). We could not apply the contextual analysis on Arabic writing because the IFN/ENIT base only provides isolated words.

Thanks to the use of the context, we obtain excellent results for baseline positioning: more than 98% of *good* (error <5 pixels) baselines on French and Bangla bases, with our generic method. The precision is really increased (more *good* lines). The figure 6 illustrates some example of beneficial use of the context for baseline positioning: it is particularly important for short words and text lines with slope. Finally, the important increase obtained for these scripts is promising for Arabic baseline extraction using the context.

Baseline quality (average pixel error)	French (RIMES project)		Bangla	
	Without context	Using global line	Without context	Using global line
Good (<5)	80.9 %	98.4 %	69.7 %	98.0 %
Acceptable (<7)	85.6 %	98.5 %	83.2 %	98.5 %

Table 4. Using the context of the global line improves baseline quality



(a) Baseline detection without context (each word is treated independently)

(b) Baseline detection using indications about the context of the text line

Figure 6. Example of baselines that are badly recognized without context (circled)

6. Conclusion

In conclusion, we have presented in this paper a new method for baseline detection. It is based on the perceptive vision, and consists in using the global properties of the word (slope, thickness) at low resolution with an accurate analysis at high resolution, in order to extract the precise position of the baseline. This method is generic because it does not take into account any specific properties of an alphabet, contrary to the approaches of the literature. Thus it has been applied on three kinds of script, without any parameter adjustment: Arabic, French, Bangla. We evaluate the obtained results using a public metric proposed by IFN/ENIT, which enables a precise comparison of our method with existing work. The results show that our generic method obtains recognition rates comparable to existing methods dedicated to Arabic writings

Moreover, we demonstrate how the use of knowledge on the initial context of the word (text line) significantly improves baseline extraction, as it provides more reliable information on the slope and the thickness. Thus, thanks to the context, we obtain more than 98% *good* baselines (only 83-85% without context).

As a conclusion, we have shown that the use of a perceptive vision mechanism provides good results for baseline extraction of several scripts, and more particularly when the global context is taken in consideration.

References

Abed, H.E. & Märgner, V. (2007). The IFN/ENIT-database - a tool to develop Arabic handwriting recognition systems *9th International Symposium on Signal Processing and Its Applications (ISSPA 2007).*

Augustin, E., Carre, M., Grosicki, E., Brodin, J., Geoffrois, E. & Preteux, F. (2006). RIMES evaluation campaign for handwritten mail processing. *International Workshop on Frontiers in Handwriting Recognition,* pp. 231-235.

Bhattacharya, U.; Parui, S.K. & Shaw, B. (2007). A hybrid scheme for recognition of Handwritten Bangla basic characters based on HMM and MLP classifiers. *International Conference on Advances in Pattern Recognition,* pp 101-106.

Cinque, L.; Lombardi, L. & Manzini, G. (1998). A multiresolution approach for page segmentation. *Pattern Recognition Letters, Elsevier Science Inc., 19,* pp 217-225.

Coüasnon, B. (2006). DMOS, a Generic Document Recognition Method: Application to Table Structure Analysis in a General and in a Specific Way. In *International Journal on Document Analysis and Recognition, 8(2),* pp 111-122 .

Déforges, O. & Barba, D. (1994). A Fast Multiresolution Text-Line and Non Text-Line Structures Extraction. *International Conference on Image Processing (ICIP),* pp 134-138.

Lemaitre, A., Camillerapp, J. & Coüasnon, B. (2008). Multiresolution Cooperation Improves Document Structure Recognition. In *International Journal on Document Analysis and Recognition (IJDAR),* pp 97-109.

Lemaitre, A., Chaudhuri, B.B. & Coüasnon, B. (2007) Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition. In *International Conference on Document Analysis and Recognition (ICDAR'07),* pp 614-618

Maddouri, S.S., Samoud, F.B., Bouriel, K., Ellouze, N. & Abed, H.E (2008). Baseline extraction: Comparison of Six Methods on FN/ENIT Database. In *International Conference on Frontiers in Handwriting Recognition (ICFHR'08).*